SSCI 583 Spatial Analysis
Project 2 Report

Group member: Junhong Duan, Lingduo Luo, Ruochen Liang, Shiqi Li, Yongchun Chen

1 Introduction

This project explores the spatial pattern of police violence across New York City using point pattern methodologies. As law enforcement increasingly incorporates GIS and spatial analyses into their day-to-day workflows, it is essential to understand how administrative demarcations of space impact the lived experience of space across cities. This project seeks to create a counter-map that pushes back against traditional crime risk maps to offer a different view of the risk of violence in urban space.

Police violence is a critical issue in the US, particularly in specific communities and certain groups of people. Unfortunately, police agencies are not required to report incidents of police violence, so there is no official, comprehensive data set on police violence. To address this knowledge gap, researchers have created crowdsourced datasets, such as the Fatal Encounters data set, to track incidents of fatal police violence in the US.

Our project aims to explore the spatial pattern of police violence across New York City using point pattern methodologies. We will use the Fatal Encounters dataset, which tracks incidents of fatal police violence in the US from 2000 to the present, to create a counter-map that offers a different view of the risk of violence in urban space.

Our methods will include visualizing the pattern of fatal police encounters with a kernel density map, assessing the point pattern of fatal police encounters using Ripley's K method, and devising an analysis that accounts for the inhomogeneity of the background environment. The project sits within the broader project of Critical GIS, and we will be working in pairs to complete both the technical work and the written product.

2 Study area

The study area for our project is New York City, located in the northeastern part of the United States. New York City comprises five boroughs - the Bronx, Brooklyn, Manhattan, Queens, and Staten Island - each of which has unique neighborhoods and demographics.

To help the reader understand the place we are studying, we have created two maps that display relevant data layers. The first map shows the population distribution across New York City, using data from the US Census Bureau's American Community Survey. The second map shows the point locations of fatal police encounters in New York City, using data from the Fatal Encounters database.

The population distribution map is a map that visualizes population density by census tract. Darker shades of blue represent areas with higher population density, while lighter shades of blue represent areas with lower population density. The map clearly shows that the population is concentrated in Manhattan and the western parts of Brooklyn and Queens, with lower population density in the Bronx, Staten Island, and eastern Queens.

The second map, showing the point locations of fatal police encounters, is a point map that displays each incident as a purple dot. The map illustrates the spatial pattern of fatal police encounters in New York City and shows that they are not evenly distributed across the city. Instead, there are clusters of incidents in certain areas, such as parts of Brooklyn and the Bronx.

Together, these two maps help us understand our study area's context and the spatial distribution of relevant variables.

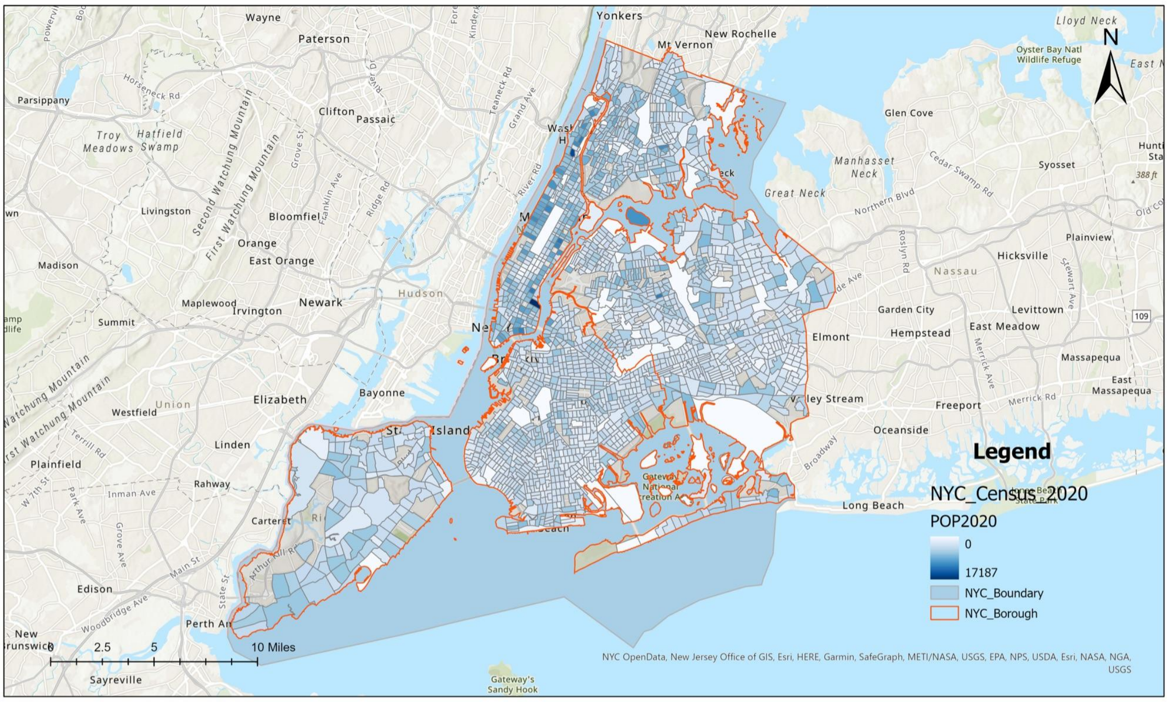

Fig. 1 Study Area and NYC Census Data Visualization

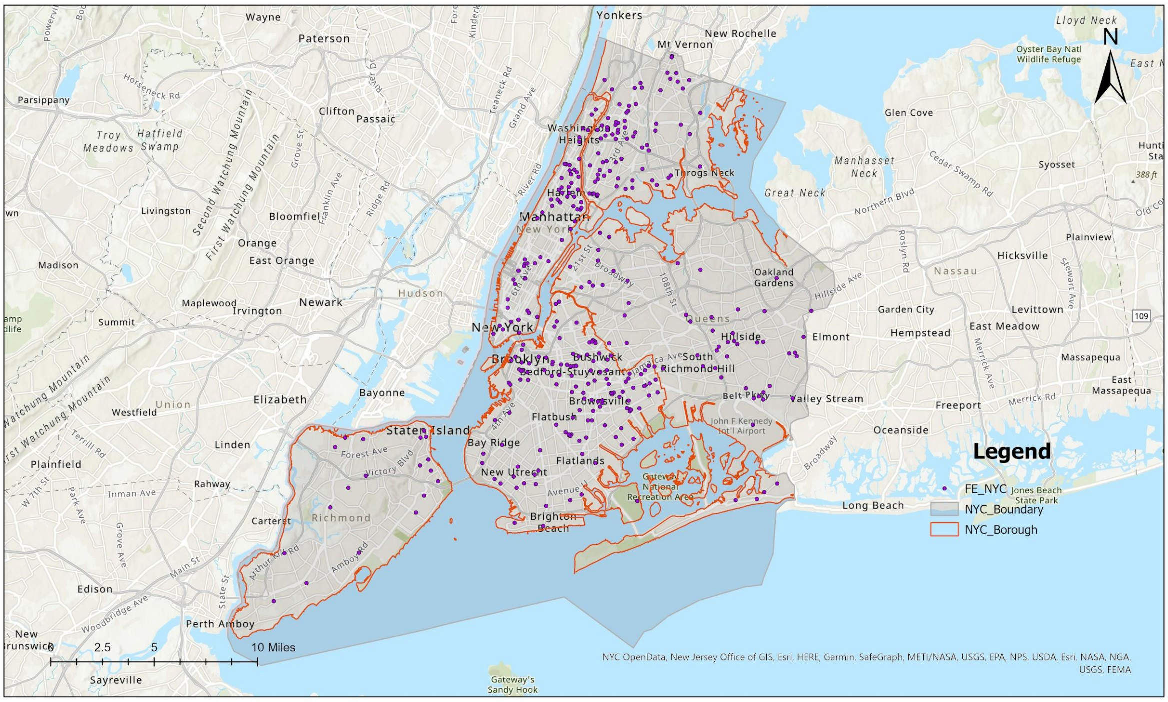

Fig. 2 Study Area and NYC Fatal Encounter Data Visualization

3 Data and Methods

3.1 Data

The Fatal Encounter Data is a dataset that collects information on fatal encounters between civilians and law enforcement officers in the United States. The data is publicly available and can be downloaded from the Fatal Encounters website. The dataset includes information such as the date and location of the encounter, the race and gender of the civilian, the reason for the encounter, and whether any weapons were involved.

This data is in a tabular format, with each row representing a single encounter and each column representing a different attribute of that encounter. We import the dataset in vector point data format with an attribute table.

Dataset Name/Description	**Data Format**	**Source of Data**	**Coordinate System**
NYC_Census_Tracts_2020	Polygon layer	NYC Open Data portal	NAD 1983
NYC_Borough_Boundaries	Polygon layer	NYC Open Data portal	NAD 1983
NYC_Boundary	Polygon layer	Created by dissolving the boundaries on the NYC Borough Boundaries data layer.	NAD 1983
FE_NYC	Point layer	Fatal Encounters website (Point layer created from the FE spreadsheet) https://fatalencounters.org	NAD 1983
EJI Dataset (Social Vulnerability)	Polygon layer (csv table joined with NYC_Census_Tracts_2020by FIPS code)	Data from the U.S. Census Bureau, the U.S. Environmental Protection Agency, the U.S. Mine Safety and Health Administration, and the U.S. Centers for Disease Control and Prevention to rank the cumulative impacts of environmental injustice on health for every census tract.	NAD 1983

Table 1 Data Table

3.2 Methods

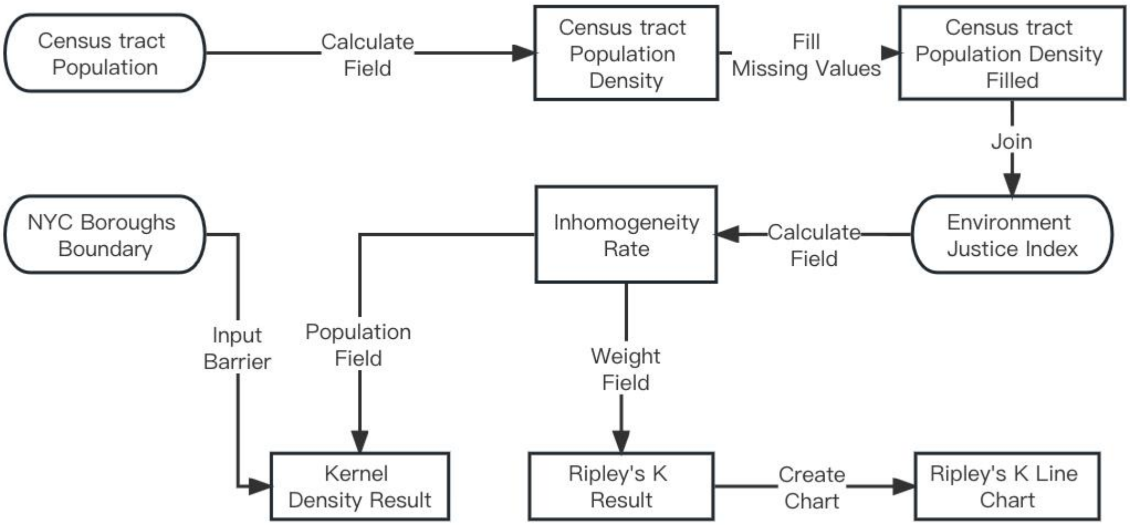

Fig. 3 Study Workflow and Methods

3.3 Data Preparation

As is listed in the data table, the data used in the study all share the same coordinate system (NAD 1983), so projection is not needed. We use the Fill missing values tool in ArcGIS Pro to fill some missing lines in the POP2020 field of NYC_Census_Tracts data, using the k-nearest neighbors method.

3.4 Inhomogeneity

To address the problem of spatial inhomogeneity in crime analysis, we propose the use of population density scores and social vulnerability scores. Social vulnerability is one component of the Environmental Justice Index (EJI) [3], which measures the cumulative impact of environmental injustice on health in each census tract, using data from various sources such as the U.S. Census Bureau, the U.S. Environmental Protection Agency, and the U.S. Centers for Disease Control and Prevention. Studies have shown that individuals from marginalized groups, such as Black individuals, are more likely to be subject to police violence and crime in areas with higher social vulnerability [4]. Accounting for population density is also important as it can impact the frequency of reported crime events and lead to biased conclusions about the spatial distribution of crime.

In this study, we will calculate the Z Score by using the Standardized Field tool to combine the social vulnerability overall score from the EJI dataset (SVM_SPL) and population density calculated by dividing the population of each census tract from the NYC_Census_Tracts_2020 dataset by its geographic area (POP2020). We will then join this data with the FE_NYC dataset using the Spatial Join tool with a search radius of 100 meters to avoid issues with low-population-density-census-tract-boundaries surrounded by higher-population-density-census-tract boundaries. This approach will help to provide a more accurate and unbiased analysis of crime in the study area.

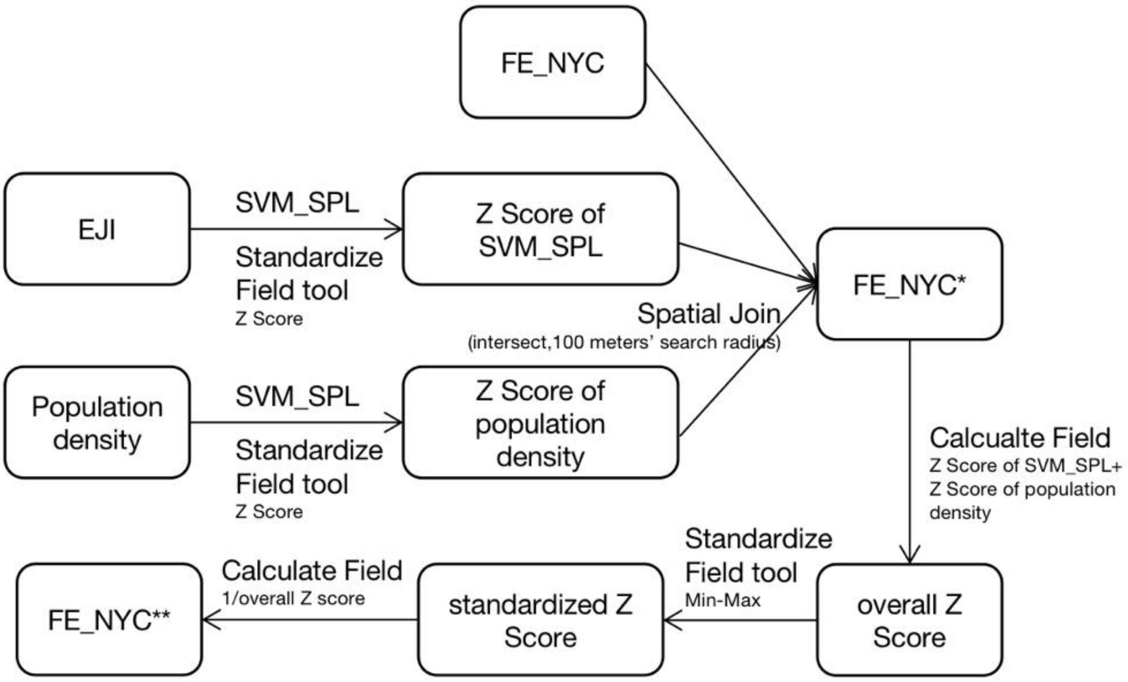

Fig. 4 Workflow of Inhomogeneity

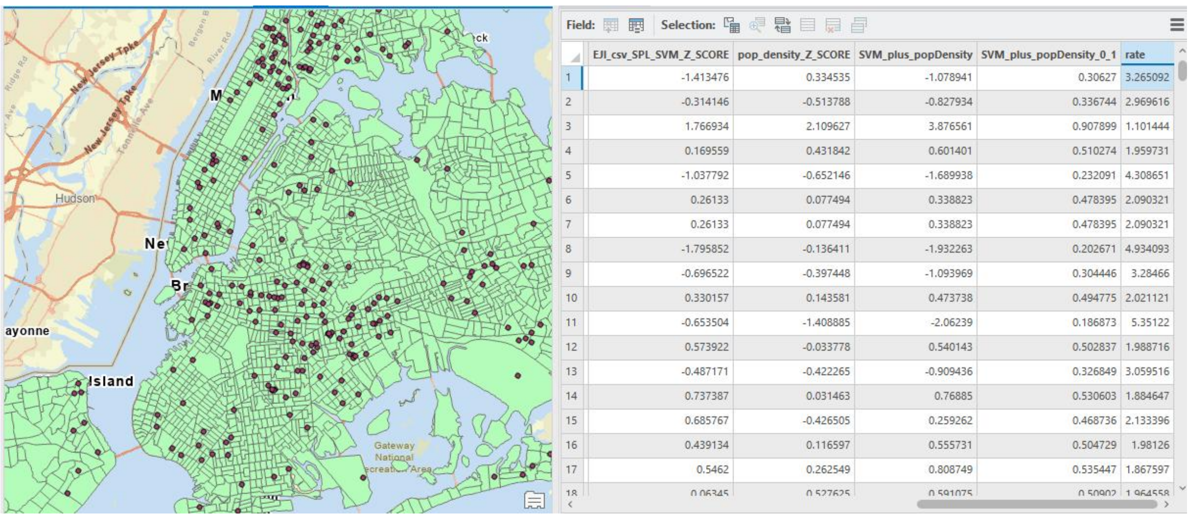

Fig. 5 Study Results of Inhomogeneity

Since higher SV and higher population density will induce higher frequency of reported FE events, we will add the two Z Scores up for each FE point. Then we will use the Standardize Field tool to transfer the sums into values between 0 and 1.

As O' Sullivan et. al [5]. talked about dealing with inhomogeneity, expressing the counts of events as some rate of incidence will work well. Therefore, we will assume each point is one event and calculate the rate as 1 divided by Z Scores summed up (named as Rate).

3.5 Kernel density and Ripley's K Methods

Kernel density is a method to create a density surface representing the distribution of Police Violence points across New York City. It deals with its inhomogeneity problem from population density and social vulnerability variation. And then, by analyzing the point dataset using Ripley's K, it is possible to identify areas with a higher density of events than would be expected by chance. Additionally, by analyzing multiple distance scales simultaneously, the multi-distance Ripley's K analysis can reveal whether the spatial clustering patterns are consistent across different scales.

Therefore, we will use the Kernel Density to create a FE density raster and use the Multi-Distance Spatial Cluster tool to identify if FE events are clustered or dispersed, and if so, at what distances.

4 Results

4.1 Kernel density

In the Kernel Density tool, we will set the Rate we calculated in the last steps as the Population Field. The output cell size is 500, which is appropriate for visualization. The input barrier features parameter is set as NYC_Borough_Boundaries, which excludes water and lakes. The reason why we use sets of continuous boroughs is that we do not expect there to be FE events that happen over water areas. Since some FE points belong to census tract boundaries with very low population density and social vulnerability outliers, it will give extremely high weight to those event points. To avoid this, we deleted the Rate value higher than one standard deviation. The final map is shown below:

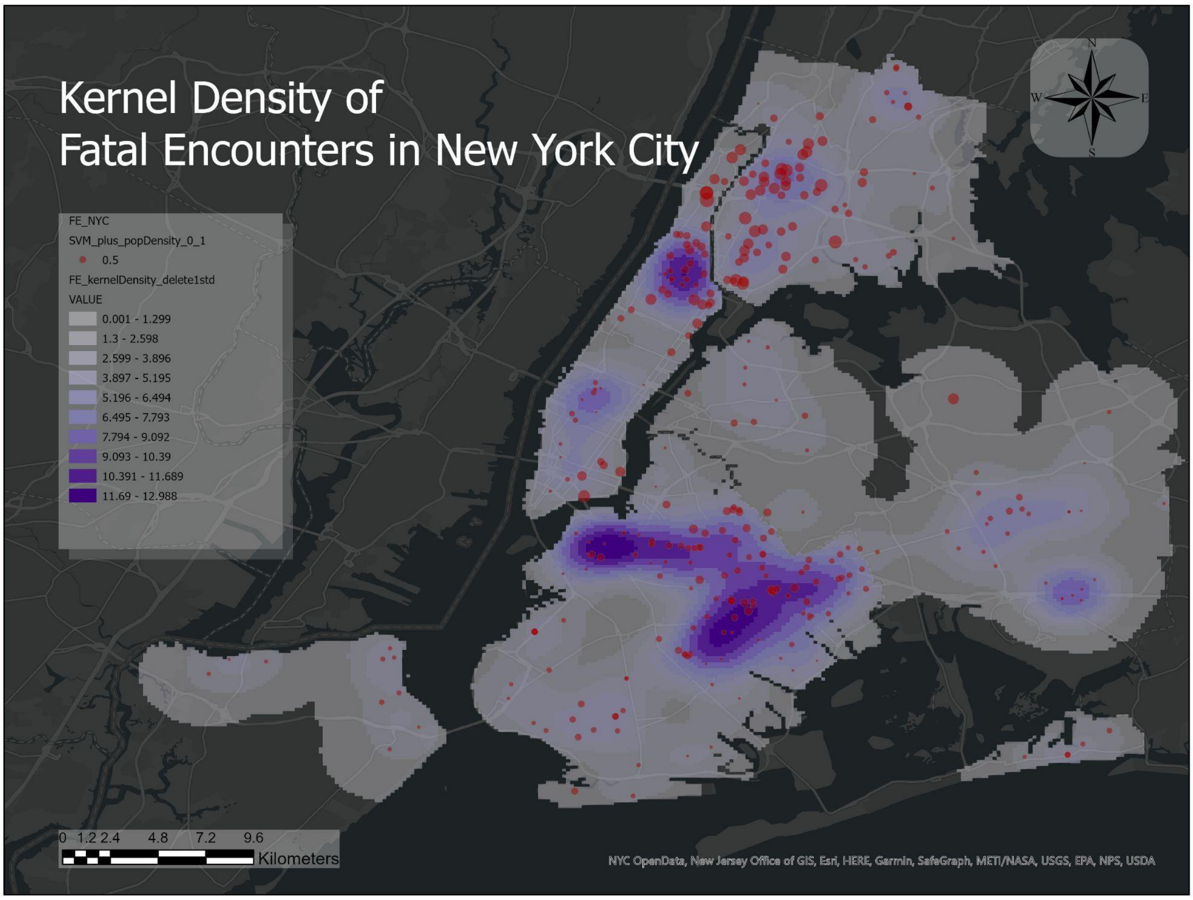

Fig. 6 Kernel Density of Fatal Encounters in New York City

In this kernel density final map, dark purple represents a high density of fatal encounters after removing inhomogeneity problems. Red dots are where the FE events happened. The larger the dot, the higher the Z Score (SV + Population density). From the plot, we can see that in Brownsville, Brooklyn, and Harlem, FE events happen very intensively. Although many FE events happened closely, for instance, in Tremont and Morrisania, the higher population density and social vulnerability values make their FE Kernel density have lower values.

4.2 Ripley's K

In this project, we use the FE_NYC after spatial join as the input feature class, set 20 distance bands, Rate as Weight Field, and use all default values for the rest of the parameters, except for the boundary correction method, we will choose to Simulate Outer Boundary Values to tackle with edge effects. The Ripley's K line chart result is shown below:

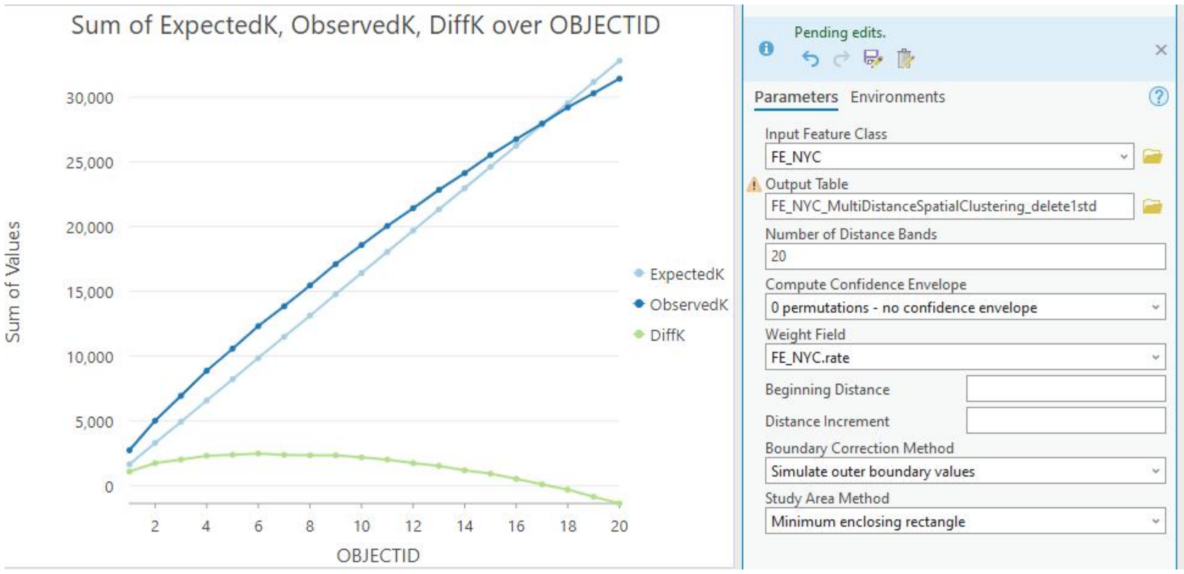

Fig. 7 Ripley's K Line Chart Result

Ripley's K method will be affected by underestimating the number of neighbors for features near the edges of the study area. Simulate outer boundary values will simulate points (mirrors) outside the study area so that the number of neighbors near edges is not underestimated. The points clustered in the ObservedK are higher than Expected K and scattered otherwise.

From the result, we can learn that FE events are clustered in a 17-unit distance (27896.5 feet). From the distance of 17 units, FE events could be considered dispersed. When K equals 6 (9845.8 Feet), FE events have the most clustered pattern.

5 Discussion

5.1 What do the results mean in the real world

In 2022, police shootings were responsible for 97% of all fatalities caused by police, with the remaining deaths attributed to Tasers, physical force, and police vehicles. [6] This makes analyzing police violence events a significant matter for public and government management. The analysis can reveal underlying factors or conditions contributing to police violence in certain areas, such as social inequality, poverty, or racial discrimination. In the real world, this clustering of police violence can seriously affect the affected communities. It can lead to a breakdown in trust between the police and the community, making it harder for police to carry out their duties effectively and for community members to feel safe and secure. It can also lead to a higher incidence of injuries, fatalities, and wrongful arrests or convictions, particularly among minority groups more likely to be targeted by police violence. [7] Therefore, understanding patterns of police violence is critical for the safety of societies and people, which can help develop prevention and intervention strategies and promote accountability and transparency in law enforcement.

The project's results provide insights into the spatial distribution of police violence in New York City, highlighting areas where violence is more likely to occur. Higher population density and greater social vulnerability correlated with higher rates of police violence. Therefore, the research must consider these factors to obtain a more accurate picture.

Considering inhomogeneity from population density and social vulnerability, this analysis offers a more precise understanding of police violence prevalence in different areas. Identification of clustering and dispersed zones can guide policy and resource allocation decisions to reduce instances of police violence. Given the severity of police violence in America, this project holds important implications for promoting police accountability and protecting citizen rights.

5.2 Limitations in results:

The study revealed a density map indicating the prevalence of fatal police violence in New York City, which accounted for the effects of population density and social vulnerability. Clustering of such violence was found in Brownsville, Brooklyn, and Harlem.

However, Ripley's K method assumes complete spatial randomness when analyzing this violence clustering. While in reality, this pattern may have some degree of spatial correlation or clustering due to underlying processes or factors, which can lead to biased results. Besides, the research did not explore other potential factors contributing to higher incidence rates of police violence in these neighborhoods, such as crime rates, racial and socioeconomic disparities, police policies and practices, and community-police relations.

Overall, given the limitations of the methods used and the reasons for the observed clustering may be complex and multifaceted, it is necessary to use other complementary methods, such as other spatial statistics, visualization techniques, or modeling approaches, in order to gain a more comprehensive understanding of the point pattern and consider more social and human factors in further analysis and research.

6 Reference

[1] Zheng, P. et al. "Edge-correction for spatial kernel smoothing methods? When is it necessary?" (2004).
[2] Centers for Disease Control and Prevention. "ATSDR: Environmental Justice." Accessed February 19, 2023. https://www.atsdr.cdc.gov/placeandhealth/eji/index.html.
[3] Arthur Charpentier, Ewen Gallic. Kernel density estimation based on Ripley's correction. Geoinfor-matica, 2016, 20 (1), pp.95-116. 10.1007/s10707-015-0232-z . halshs-01238499
[4] Edwards, F., Lee, H., & Esposito, M. (2019). Risk of being killed by police use of force in the United States by age, race-ethnicity, and sex. American Journal of Public Health, 109(6), 941-946. doi:10.2105/AJPH.2018.304916
[5] O'Sullivan, D., & Unwin, D. J. (2010). Geographic information analysis. John Wiley & Sons.
[6] Police Violence Report. (2021). 2021 Police Violence Report. Policeviolencereport.org. https://policeviolencereport.org/
[7] Braga A A, Papachristos A V, Hureau D M. The effects of hot spots policing on crime: An updated systematic review and meta-analysis[J]. Justice Quarterly, 2014, 31(4): 633-663.

7 Appendix

Junhong Duan: Discussion and Practice: Inhomogeneity, Ripley's K, Kernel density, edge correction. Written Part 3,4,5
Lingduo Luo: Discussion and Practice: Inhomogeneity, edge correction, Ripley's K, Kernel density. Written Part 1,2,3,5
Ruochen Liang: Discussion and Practice: Inhomogeneity, Kernel density, Ripley's K. Written Part 3,4,5
Shiqi Li: Discussion and Practice: Inhomogeneity, Kernel density, Ripley's K, edge correction. Written Part 3,4,5
Yongchun Chen: Discussion and Practice: Inhomogeneity, Kernel density, Ripley's K. Written Part 4,5